BEST PRACTICES IN CROSS-DEVICE AND CROSS-CHANNEL IDENTITY MEASUREMENT

Advisor:
Evan Neufeld, storyline development

September 2016

EXECUTIVE SUMMARY

Through developments in new technology, the advertising and marketing industry has been moving closer and closer to what has been called "People-Based Marketing." Although it has been discussed for many years, the practice of mapping consumer identities across all channels (online and offline) and digital devices (PCs, mobile phones, tablets, smart TVs, etc.) has been, until recently, largely focused on matching personal computer (PC) data to offline data and performed by a select number of brands.

In the past 18 months this practice has seen a dynamic rise in both importance and adoption and has been driven by the explosion in consumers' use of mobile devices. This rapid expansion of digital media and other consumer touch points poses both opportunities and challenges to advertisers. It offers them the opportunity to reach consumers in a highly holistic, targeted, interactive mode (often referred to as **Identity-Based** or **People-Based Marketing**), and challenges them to connect with consumers and measure the impact of their advertising across the various touch points, while also raising questions about how to respect consumer privacy.

The Coalition for Innovative Media Measurement (CIMM) commissioned this white paper to better understand the census-based methods for establishing consumer identity across multiple platforms and Internet-connected digital devices (including smart TVs) as well as methods for linking offline and online behavior for the same household or individual via links between home mailing addresses and IP addresses, e-mail addresses, mobile phone numbers, landlines, device IDs, cookies, and so on.

The goal of this white paper is to provide the state of the art in "identity" technologies, describe the landscape of different companies involved, and identify any areas for further innovation by CIMM. To facilitate this objective, CIMM identified and conducted in-depth interviews with 20 key participants in the ecosystem over a three-month period from July to September 2016. These interviews allowed CIMM to highlight best and worst practices currently in the market as well as provide useful frameworks and guidelines that brands could apply when conducting cross-device/cross-platform exercises in consumer matching. Key Findings of this report include:

- Matching consumers across devices and platforms is currently top of mind with US marketers. In a recent survey, nearly 70% of US digital marketers and media practitioners cited cross-device audience recognition as a topic that would command most of their attention in 2016.

- For the foreseeable future, most marketers will have to work with a variety of solutions deploying deterministic, probabilistic, and (increasingly) hybrid approaches to create "best-of-breed" solutions to combine scale with accuracy.
- Truth File access and data points on which solution providers match are the key variables influencing **match rates**.
- Key measures to understanding the true implication of match rates are **accuracy** and **precision**, followed by **reach** and **scale**.
- Beyond simply focusing on overall match rates, marketers must ensure they understand the individual data points that contribute to those matches and their respective strengths and weaknesses.
- For marketers to maximize and better understand how they can increase the impact of any identity-based marketing campaign, they need to have a firm grasp on how providers weigh and balance data in their various graphs, and they need to know how to measure the ability of those graphs to target the correct consumers across devices and platforms.
- Ad tech companies that provide services to help match consumers across devices present marketers with a confusing, overlapping ecosystem of offerings to navigate. This appears unlikely to change over the next 12 to 18 months.
- The intersection of increased government scrutiny and heightened consumer concerns around privacy requires marketers approaching identity-based marketing to have a deep understanding of the data sources they are working with and how they do or do not protect consumer privacy.
- Areas that require additional focus by and across industry bodies include: industry-level education in terms of best practices for identity matching and assessment of data quality, stronger clarity and standards about protecting consumer privacy, and industry standardization around definitions of key measures and agreement on how they should be calculated.

Table of Contents

INTRODUCTION:

The Rise of Identity-Based or People-Based Marketing

Although it has been discussed for many years, the practice of mapping consumer identities across all channels (online and offline) and devices (PCs, mobile phones, tablets, smart TVs, etc.) has been, until recently, largely focused on matching personal computer (PC) data to offline data and performed by a select number of brands.

In the past 18 months this practice has seen a dynamic rise in both importance and adoption, driven by the explosion in consumers' use of mobile devices. **According to an Ericsson Research study, the average US household possesses 5.2 Internet-connected devices.**[1] Nielsen places this number at an average of four digital devices and states that users in the United States engage with media content across screens for more than 60 hours per week.[2]

Concurrent with their usage of more devices to interact with content, consumers have increased expectations regarding the sophistication of cross-channel brand messaging. **According to data from Forrester Research, seven in ten consumers have a negative view of inconsistent cross-channel messaging, and one in ten feel that inconsistencies in brand experiences would stop them from interacting with a brand altogether.**[3]

This rapid explosion of digital media and other consumer touch points poses both opportunities and challenges to advertisers. It offers them the opportunity to reach consumers in a highly targeted, interactive mode and challenges them on how to measure the impact of advertising across the various touch points.

Tracking consumers across multiple platforms and Internet-connected digital devices (including smart TVs), linking offline and online behavior for the same household or individual via links between home mailing, IP, and e-mail addresses, mobile phone numbers, landlines, device IDs, and cookies now appears to be a mandatory practice for constructing and executing advertising strategies. **In a recent survey, nearly 70 percent of US digital marketers and media practitioners cited cross-device audience recognition as a topic that would command most of their attention in 2016.**[4] More importantly, interest in this practice, often referred to as identity-based marketing, saw the greatest year-over-year jump of any mentioned tactic. "The ability to accurately target and track consumers across devices ('cross-device') has become 2016's top make-or-break marketing competency," noted Richard Joyce of Forrester Research in his recent report, "Bridging the Cross-Device Chasm."

Not surprisingly, as with many nascent digital advertising practices, executing identity-based marketing is incredibly complex and difficult on several levels. On the infrastructure side, over the last 12 to 18 months, there has been explosive growth in the

---

[1] "Understanding Today's Cross-Device Consumer," DataXu, February 2016.
[2] "The US Digital Consumer Report," Nielsen, October 2015.
[3] "Bridging the Cross-Device Chasm," Forrester. November 2015.
[4] "The Outlook for Data: 2016 Snapshot," Winterberry Group, January 2016.
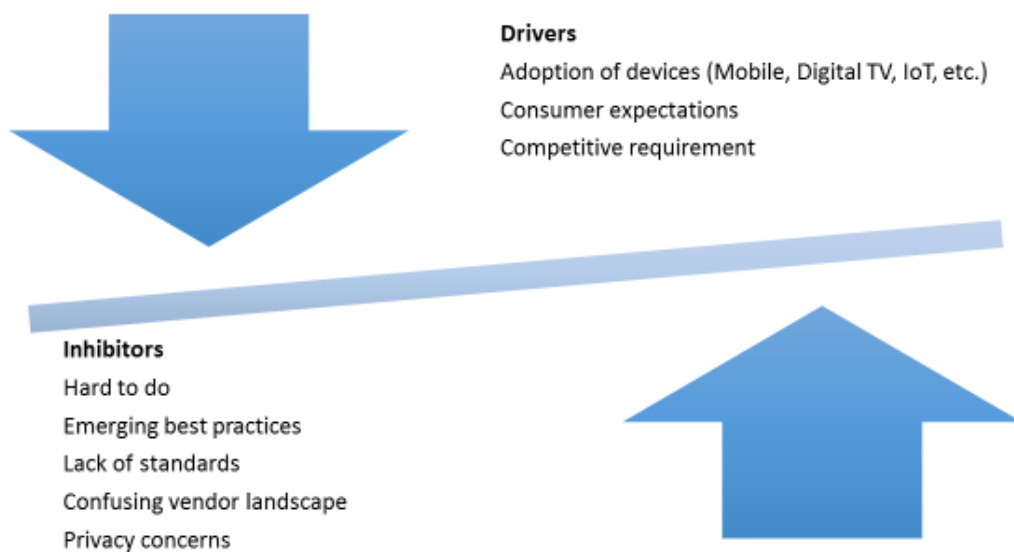
breadth and depth of cross-device audience identification capabilities, including audience measurement companies such as Nielsen and comScore. Both Facebook and Google are in possession of known cross-device solutions that are capable of approximating identities using publicly available data and machine learning.

Although there is no shortage of cross-device options and approaches, brands remain limited in their ability to apply a single cross-device lens across every platform, property, and channel. Cross-device targeting "came to the masses in 2015," according to Alan Beiagi, senior director of products at DSP DataXu. "That resulted in a lot of conversations about deterministic and probabilistic data," he said. "But what we still see is that cross-device attribution is still very much in its infancy, and there are no standards around it yet. Between what ID you have on one platform versus what kind of ID you have on another, it ends up becoming very complicated. . . and the resulting attribution can look entirely different."

However, as has historically been the case in digital marketing technology, the sophistication of both best practices and tools that are currently available in the space is lagging significantly behind demand. "To describe the space as evolving would be a vast understatement," noted one marketing executive we talked to, adding, "brands looking to deploy turnkey, cookie-cutter approaches here will find themselves at a loss."

"At this point in the game, it's safe to say that every sophisticated marketer understands the challenges inherent with the old way of doing things vis-à-vis last-touch modeling, cookie-based solutions, and so forth," noted Ari Buchalter, President of Technology at MediaMath. "I think we're on to a stage where everyone agrees on the problem and is trying out a series of solutions but is running up against the limitations of those solutions and are trying to figure out how to deal with that."

1 - Drivers & Inhibitors of Identity-Based Marketing

Drivers
Adoption of devices (Mobile, Digital TV, IoT, etc.)
Consumer expectations
Competitive requirement

Inhibitors
Hard to do
Emerging best practices
Lack of standards
Confusing vendor landscape
Privacy concerns

Approach:

The Coalition for Innovative Media Measurement (CIMM) commissioned this white paper to better understand the census-based methods for establishing consumer identity across multiple platforms and Internet-connected digital devices (including smart TVs) as well as methods for linking offline and online behavior for the same household or individual via links between home mailing addresses, IP addresses, e-mail addresses, mobile phone numbers, landlines, device IDs, cookies, and so on.

The goal of this white paper is to provide the state of the art in these "identity" technologies, describe the landscape of different companies involved, and identify any areas for further innovation by CIMM. To facilitate this objective, CIMM identified and conducted in-depth interviews with 20 key participants in this ecosystem over a three-month period from July to September 2016. These interviews allowed CIMM to highlight best practices currently in the market and provide useful frameworks and guidelines that brands could apply when conducting cross-device/cross-platform exercises in consumer matching.

I. Key Finding: Deterministic and Probabilistic Data Sets

Currently, when discussing the process of matching consumer identities across devices and platforms, solutions have been broadly characterized as being either deterministic or probabilistic in nature.

Deterministic—Scaling the Walled Garden

Deterministic data, often referred to as first-party data, originates from a company, such as Facebook, Google, or Verizon/AOL, that has the ability to leverage links across devices with a 100% level of confidence using login or subscriber data. For example, a user will log in to a streaming video service on his or her mobile app with the same log-in information he or she uses on his or her desktop or will sign into his or her Facebook account on a computer and then on a mobile phone and be identified as the same user. Although deterministic data sounds like a panacea, it has several well-known issues, including the following:

1. **Accuracy**: Deterministic data provides extremely strong indicators, but in practice is not always 100% accurate because of factors that include cookie deletion, multiple users accessing content from a single device, and/or multiple users of log-in credentials across different devices. In fact, comScore's validation of AOL's deterministic data set pegged its accuracy at 93%.[5]
2. **Privacy**: Because deterministic data sets contain personally identifiable information (PII), their use raises privacy concerns. We will explore the complex issues around consumer privacy in detail in Section VIII of this report.

---

[5] "AOL Cross-Screen Validation Report," comScore, October 2015.

3. **Walled gardens**: Although first-party data providers, such as Google, Facebook, and Amazon, have scalable log-in data from their massive user bases, they are not always good about sharing such data or they may be restricted in their ability to share data because of privacy policies or contractual commitments, resulting in virtual "walled gardens" when it comes to identifying users across screens.

4. **Limited scale**: Although highly accurate, this type of data is limited in scale, especially as the target audience becomes more specific and defined by more and more attributes.

As one agency executive noted, "Doing identity-based marketing using deterministic data is great for broad-based targets focused on, say, gender or age, but as the target gets more sophisticated, the limits of deterministic data become more apparent. Additionally, the "black-box" aspect of working with first-party data can leave you feeling a bit blind to what is actually going on. Targets are set, campaigns are run, but while you see results, you are much less able to access and understand the process by which those initial matches are made and fine-tuned."

Probabilistic—Reaching the Masses

The probabilistic data approach is built around the aggregation of scores of data points, such as behavioral data or data that may indirectly identify a specific individual (e.g., device identifiers and cookies). Although this data on its own may not readily identify a specific individual, in many non-U.S. countries such data may still be considered PII or "personal data." And, as discussed in section VIII of this report, regulators in the United States may increasingly share this view. For purposes of clarity in this paper, we refer to such data as "indirectly identifiable data" or "II Data." While II Data is less sensitive than PII, II Data sets may still raise privacy concerns. Such II Data is processed to determine the strongest probability that a set of devices are related. From here, algorithms further process the data, seeking confirmatory usage patterns that these devices have been properly matched. II Data analyzed in this context can include location (multiple devices always seen with each other via an identical Internet connection are probably related); content usage (comparing the types of sites and content that are consumed on various devices; and time (analyzing device wake/sleep patterns).

The real value of probabilistic modeling is the ability to scale. By not being limited to any single walled garden or group of people logging in to a common platform (such as social media or e-mail) and focusing instead on making sense of a mass of anatomized data, this approach allows for a much larger pool of devices to be targeted. A disadvantage is that such approaches are only as good as the algorithms they deploy to match devices and usage patterns, making the accuracy of such matches harder to prove.

"Probabilistic approaches offer marketers solutions that scale," noted one data provider. "Typically, your tolerance for using probabilistic data really should vary based on your objective. For very broad campaigns, large geographic areas, or consumers who use a specific type of content, probabilistic approaches result in matches that, while perhaps not always perfect, are often good enough. But the fact is that these solutions have their own 'black box' in their algorithms, making accuracy of matches harder to determine.

Understanding in detail how matches are made using probabilistic approaches is absolutely key to ensuring you maximize the ROI from such solutions."

2 - Deterministic and Probabilistic Data Sets

	Deterministic	Probabilistic
Description	Solutions that stitch together data based on login or other consumer provided data.	Solutions that predict common device ownership by analyzing large amount of consumer internet usage data
Pros	• Mass Reach	• Niche Targeting • Non-PII
Cons	• Walled garden • Scale • PII	• Often hard to validate
Vendors	Facebook, Google, Yahoo, Verizon/AOL, Twitter, etc.	TapAd, Drawbridge, Screen6, Crosswise, etc.

Hybrid Approaches: The Best of Both Worlds?

Increasingly, the distinction between purely deterministic and purely probabilistic approaches has blurred, as several probabilistic vendors acknowledge that they use some first-party data at several points in their matching process and vice versa. For example, Acxiom-owned LiveRamp, while relying on **deterministic** data to power its Customer Link product, leverages **probabilistic** matches for clients who want additional reach, using Drawbridge's Connected Consumer Graph. And **probabilistic** provider TapAd leverages a small core of **deterministic** data that it accesses via direct partnerships to train its probabilistic device graph.

"As the market demands more accuracy and reach, we are looking to incorporate the widest range of data that we can to ensure that our approach is proving the best matching possible, and first-party data is just one example of that," said one product manager at a major probabilistic provider. She added, "As long as those data are used properly, in a 100% privacy-compliant way, such a mix will allow our clients to have their cake and eat it too, eliminating the need to ever have to trade off scale for accuracy."

The Bottom Line—Working with Deterministic and Probabilistic Vendors

For the foreseeable future, most marketers will have to work with a variety of solutions deploying deterministic, probabilistic, and (increasingly) hybrid approaches. To do so as effectively as possible, marketers must take into account several factors, including the following:

1. **Check the state of your own data/target**: Not all marketers' customer data is created equal. Approaches to identity-based marketing will vary for marketers looking to target device users versus those trying to build up their own internal

data sets on customers. If the collected customer data are limited, a good first step is to seek providers that can share their deterministic data to help triangulate people and devices. For example, a publisher may have authenticated device connections that can be accessed by matching a marketer's e-mail list to a publisher's subscription information.

2. **Gauge internal bandwidth**: As Section VII of this report details, engaging in identity-based marketing is an arduous process that requires both internal alignment and significant resources, so understanding how much time and how many resources to commit to such efforts should be key in determining the extent to which you engage with vendors.

3. **Understand trade-offs**: Identity-based marketing today is a series of trade-offs between two primary factors: quality and quantity. Understand what your limitations are when working with walled gardens in terms of the audience they can provide. With probabilistic solutions, understanding the accuracy of algorithms and the scale of the data helps to determine efficacy.

II. Key Finding: Understanding Graphs

For marketers to maximize and better understand how they can increase the impact of any identity-based marketing campaign, they need to have a firm grasp of how providers weigh and balance data in their various graphs, and they need to know how to measure the ability of those graphs to target the correct consumers across devices and platforms.

The foundational element to all deterministic and probabilistic data is the use of graphs. Pioneered by Facebook, a graph simply charts the connections between various devices, attempting to match devices to either households or individuals. Successful graphs are able to find a high percentage of matches (reach) with a level of statistically relevant precision (accuracy).

Graphs depict the various connections between devices and users by deploying machine-based, algorithms that process a variety of data, including ad requests, cookies, access patterns, and behavioral segments. Probabilistic graphs rely exclusively on nonpermanent, user-resettable data, such as cookies or device IDs. Although at a basic level all graphs are designed to do the same thing, each company's graphs are considered intellectual property, with the exact science and ingredients often being intentionally obscured.

To give some idea of the scope of the data gathered and the complexity inherent in using graphs, Drawbridge, for example, has a probabilistic graphical model that makes predictions about consumers and their device ownership, and a separate Connected Consumer Graph™ that is made up of interconnected, individual device graphs, each of which consists of data stored in conditional probability tables that determine the probability that the device belongs to an individual user.

Drawbridge's construction of graphs has involved the processing of well *over three trillion data attributes*, including browser cookies, mobile device IDs, time, and behavior

data, gathered from over 50 different data partners, including mobile and desktop exchanges, advertisers, publishers, and data management platforms. From this mountain of data, Drawbridge produces device and user connection information for one billion consumers connected to more than three billion devices globally.

3 - Device Graphs - An Example
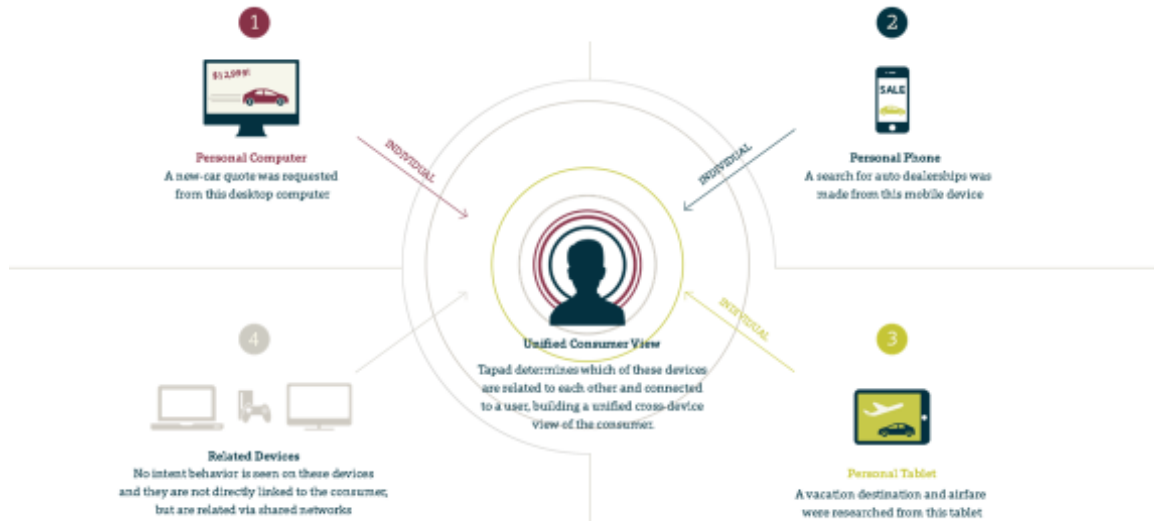
A **Device Graph** describes how different devices relate to each other by mapping of all the devices, IDs, and associated data back to one unique user or household.

Compliments of TapAd

The Bottom Line—Working with Graphs

Understanding how any vendor's graphs work is key to having confidence about their claims regarding the strength of connections across consumers and devices. Key questions that any vendor should be able to answer with regard to its graphs include the following:

1. What is the graph's match rate (specifically, to your target audience)?
2. How frequently is the rate calculated (i.e., weekly, monthly, quarterly)?
3. How do you calculate the accuracy of such matches?
4. What data goes into the graph, and is it overly dependent on any specific type of data, such as cookies or e-mail addresses?

III. Key Finding: Measures of Success

It is not surprising, given the complexity of the processes by which data are matched to a particular consumer, that definitions of success lack consistency in the marketplace across vendors.

The Holy Trinity—Match Rates, Accuracy, and Precision

The ability of a marketer to accurately match a consumer to a set of devices and connect digital profile data to offline CRM files is the core concept that drives the conversations around match rates. Overwhelmingly, the most popular measure discussed in the context of cross-platform/cross-channel consumer measurement is the ability of a vendor to match consumers. There are two types of match rates to consider.

One type of match rate is a vendor's ability to match offline data (such as Customer Relationship Management (CRM) or subscriber information) with digital data, such as log-in information, cookies, or device IDs, which is called an *onboarding match rate*. For example, let's say that Marketer X has three million e-mail addresses. With an onboarding match rate of 40% (which according to our research, is about average for onboarding data), Marketer X can expect their online audience, the number of users you can target with online display, to be around 1.2 million.

The second set of *vendor or graph match rates* reports on a vendor's ability to correctly match one set of cookie IDs to another, which in turn allows that vendor to match multiple devices to a single user.

Compounding industry confusion about match rates is the fact that several deterministic and probabilistic vendors have touted extremely high (90-plus%) match rates produced in partnership with third-party vendors such as Nielsen or comScore. As one product manager at a first-party data provider said, "Match rates are just about the most ill-used term in this space right now. To begin with, there is simply no such thing as a standard or even average match rate. These rates are highly specific to individual efforts and impacted by the data the client brings to the table as well as the complexity of the target and the accuracy of the individual graphs involved: Could I show an on-boarding match rate of 90% for females in the United States? Sure, but what would that match rate look like for 18-25 years old, who I want to target on their smartphones and via digital set top boxes? Obviously, much lower, but this should come as no surprise, because the incident of rates of this group should be equally small. More troubling still is the fact that match rates are typically a measure in time and do not account for key variables such as expired identifiers or users that opt out of tracking data over time."

Key measures to understanding the true implication of match rates are **accuracy** and **precision.**

Accuracy, in the context of cross-device identification, is typically calculated as the number of **matches** and **non-matches** correctly identified, or the number of times a deterministic or probabilistic prediction was correct. Accuracy scores vary based on approach, and how this measure is calculated varies from solution to solution. As a rule, most accuracy scores do factor in non-matching predictions to calculate accuracy. "The fact that numbers are being put out there that, in practice, count useful and non-useful matches as being equally valued is not helpful," admitted one solution provider we talked to, adding, "of course, any solution's ability to separate out complete from incomplete matches is an important metric. The way this measure is used in the marketplace is, I

think, simply confusing in the best light. This is why measuring the precision of the matches becomes so important."

Whereas **accuracy** measures the number of correctly identified matches and non-matches out of the total predicted, **precision** estimates how well the model correctly identifies **matches** only, taking into account that some of those matches will be in error.

4 - Measure of Success:
The Holy Trinity – Match Rates, Accuracy, and Precision

MATCHES

Data Onboarding Match Rates

Match offline data (such as CRM or subscriber information) with digital data, such as log-in information, cookies, or device IDs.

Device Match Rates

Reports on a vendor's ability to correctly match one set of cookie IDs to another, which in turn allows that vendor to match multiple devices to a single user.

KEY MEASURES

Accuracy

A measure of correctly identified **matches** and **non-matches**, out of the total predicted.

Precision

Estimates how well the model correctly identifies **matches**, taking into account that some of those matches will be in **error**.
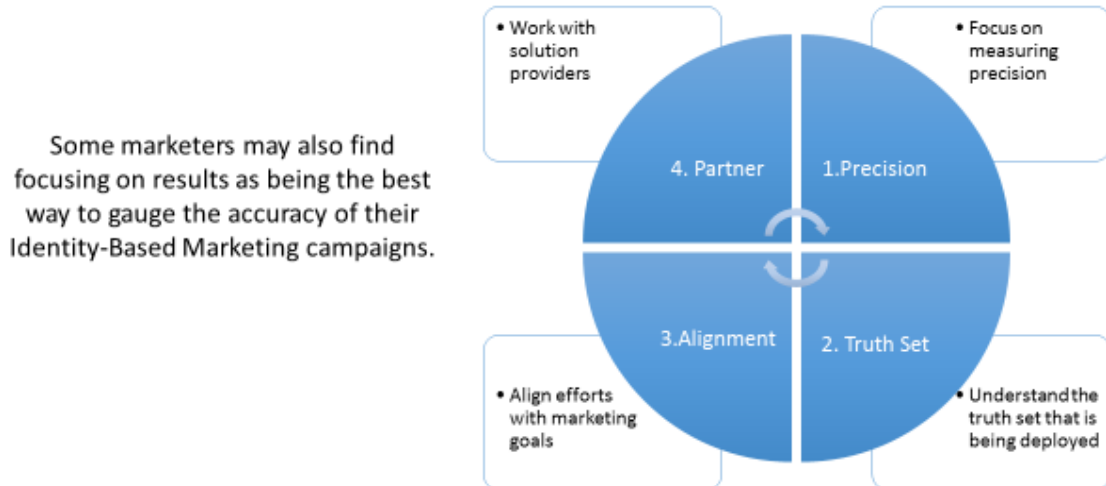
The Bottom Line—Precision Is the Key Measure

Although it will be awhile until definitions for accuracy and precision are standardized and uniformly deployed, it is important that marketers who engage in cross-device/cross-platform targeting understand the vocabulary used by vendors in the space. Conceptual familiarity with these two measures will provide more clarity into how terms are defined and can compare across vendors and the data they supply to inform execution of cross-platform campaigns. In the interim, marketers should do the following:

1. **Focus on measuring precision**: The current accuracy paradox for match rates strongly suggests that precision is a better-suited metric for evaluating the effectiveness of various approaches.
2. **Understand the truth set that is being deployed**: To the extent possible, it is more important to understand the data that are being used to create accuracy scores than the scores themselves. For example, a truth set composed of statistical IDs is less valuable than an authenticated panel truth set at measuring accuracy.
3. **Align efforts with marketing goals**: Ensure that matching efforts align with goals. For example, a marketer might trade off precision for recall (i.e., sacrificing the match being exactly right/relevant for a model that returns many matches).
4. **Work with solution providers**: Marketers must work with solution providers to make smart trade-offs on the road to matching and targeting consumers. For example, a marketer can dial down precision to achieve greater scale match rates by using looser criteria (such as a name and city versus an exact address,

matching by zip code, etc.) A supermarket, for example, may find this to be an acceptable trade off. For marketers trying to reach a specific customer, however, sacrificing precision for scale is a bad trade-off.

5 - Precision Is the Key Measure

Some marketers may also find focusing on results as being the best way to gauge the accuracy of their Identity-Based Marketing campaigns.

- Work with solution providers
- Focus on measuring precision

4. Partner   1.Precision

3.Alignment   2. Truth Set

- Align efforts with marketing goals
- Understand the truth set that is being deployed

IV. Key Finding: Macro Trends Impeding Progress

Marketers must also be aware of macro-level issues that can potentially pose obstacles to achieving further gains. Some of the barriers currently being encountered are discussed in the following list:

- **Incomplete population representation**: Audience or transactional data can be gathered from a variety of sources that may not represent the population. For example, loyalty card transactions do not capture other credit card activity or items paid for by cash. On the media side, restrictions on installing monitoring equipment on business PCs, laptops, smartphones, and tablets hamper the ability to accurately track business Internet usage. This means that a large portion of nonbusiness, daytime Internet usage is not tracked and reported by research panel companies.
- **Small sample size**: The intersection of all media exposure combined often results in a much smaller number of homes or people that are active across all media platforms. This is especially true for narrow consumer targets, such as people ages 12-24 or with $150K+ household incomes. The more precise the target demographic, the greater the challenge of reporting consumers' media or transactional activity.
- **Data integration considerations**
  - Limited matching points: Match keys between two or more data sources can be limited depending on the linking method, whether that be e-mail address, cookies, or residential address. Often, the size of the data set that is common to two sources, for example, may be 30-40% of the two original sources.
  - Integration techniques: The matching of media data with consumer demographics and transactions may not be uniform, and this can create challenges when comparing results from two different data matching sources.

- **Variance in data quality**: Data quality is one of the most significant—yet least discussed—considerations in the preparation of cross-platform measurement and reporting. Depending on the source, there will likely be differences in freshness of data, missing data, representativeness, and continuity from one reporting period to the next.

V. Key Finding: Not All Data Points Are Created Equal

Beyond focusing simply on overall match rates, marketers must ensure they understand the strengths and weaknesses of the individual data points that contribute to those matches. Factors such as joining of data sets at the household level versus the individual level as well as the look-back window for any offline data can have tremendous impact.

"There's a lot of heterogeneity in the methodologies that trip people up in terms of the final match rates," Ari Buchalter of MediaMath notes. "Some of the data are also just perpetually inaccurate. People move. They create new log-ins. There's a percentage of those data that are always inaccurate and out-of-date."

In our interviews, the following types of data points emerged as those most commonly utilized in matching. They are listed below, as are their relative strengths and weaknesses.

1. Cookies
   a. Pros: Ubiquitous on PCs; useful for building probabilistic profiles.
   b. Cons: Expire frequently; do not always work in mobile environments.
2. Internet protocol (IP) address
   a. Pros: Can be a good proxy for location; not generally considered PII by itself unless linked to other PII or, in some cases, II Data.
   b. Cons: Has no out-of-the-box correlation with a specific geographical location; some IP addresses change physical locations each month. Since the IP address is a static number and not based on a phone number assigned a geographical area code, the location of an IP address can be virtually anywhere. The more granular the Designated Marketing Area (DMA), the less reliable IP targeting can be. It is most useful for household-level matching.
3. Mobile device ID
   a. Pros: Persistent ID is accurate at identifying individuals indirectly (II Data).
   b. Cons: Does not transfer from devices when an upgrade occurs.
4. Mobile phone number
   a. Pros: Because most consumers keep their mobile numbers regardless of turnover in devices or carriers, it is an extremely stable identifier.
   b. Cons: Considered PII. Also not prevalent for some demographic groups (such as senior citizens).
5. Landline phone number
   a. Pros: Traditionally a very stable longitudinal identifier.

    b. Cons: Considered PII. Because of "cord cutting" it is much less ubiquitous among certain key cohorts, especially by consumers born after 1964); most useful for household-level matching.

6. Residential address
    a. Pros: A relatively stable identifier.
    b. Cons: Considered PII. Most useful for household-level matching.
7. E-mail addresses
    a. Pros: Ubiquitous across platforms.
    b. Cons: Consumers frequently use different e-mail addresses for different activities, so can often be hard to match; considered PII.
8. Offline purchase behavior
    a. Pros: Useful for segmentation as well as an indicator of future behavior.
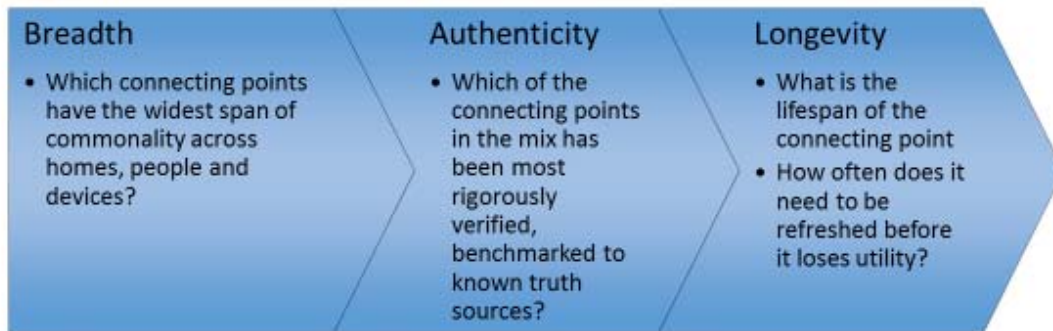   b. Cons: Hard to come by/match.

The Bottom Line—Understanding Key Data Types and Characteristics

Myriad connecting points are used in a variety of combinations during the matching process, and there is no single correct schema for establishing connecting points; however, there are points to keep in mind about the strengths and weaknesses of these connection "nodes:"

1. **Breadth:** Which connecting points have the widest span of commonality across homes, people, and devices?
2. **Authenticity:** Which of the connecting points in the mix have been most rigorously verified and benchmarked to known truth sources? Are the nodes arrived at through deterministic or probabilistic means?
3. **Longevity:** What is the lifespan of the connecting point? How often does it need to be refreshed before it loses utility? For example, cookies and/or device IDs expire, and some data suppliers don't immediately remove them. And, as data cleansing practices vary company by company, not every truth set is going to say the same thing.
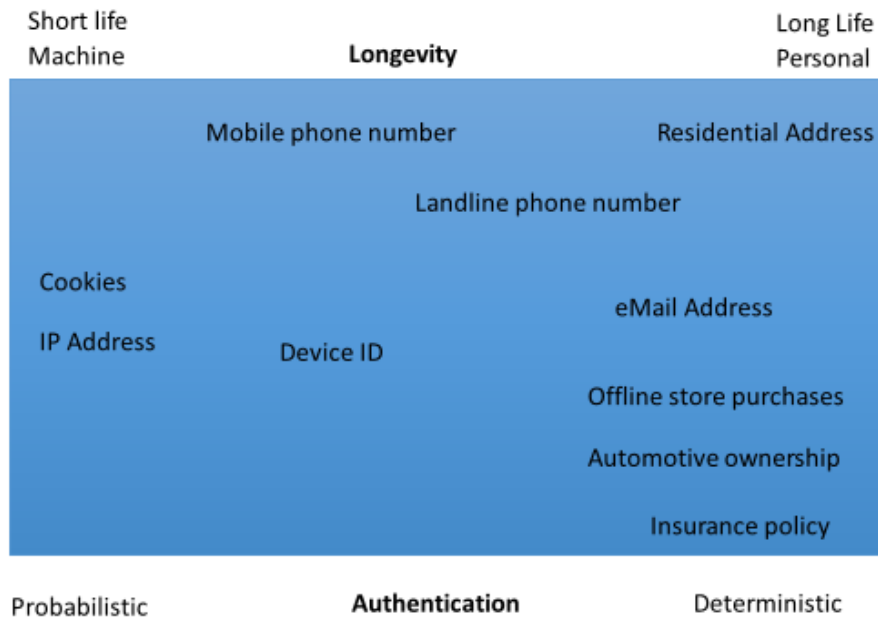
6a - Not All Data is Created Equal
Data Matching Connector Points – Evaluation Criteria

Breadth
- Which connecting points have the widest span of commonality across homes, people and devices?

Authenticity
- Which of the connecting points in the mix has been most rigorously verified, benchmarked to known truth sources?

Longevity
- What is the lifespan of the connecting point
- How often does it need to be refreshed before it loses utility?

6b - Not All Data is Created Equal
Data Matching Connector Points – Application Map

Short life Machine — **Longevity** — Long Life Personal

Mobile phone number   Residential Address

Landline phone number

Cookies

IP Address   Device ID   eMail Address

Offline store purchases

Automotive ownership

Insurance policy

Probabilistic   **Authentication**   Deterministic

VI. Key Finding: In Vendor Selection, Caveat Emptor

"While a multitude of tools present cross-device targeting and measurement as table stakes and most vendors claim to be able to bridge the cross-device gap, few can actually do this well."
- Richard Joyce, Analyst, Forrester

The number and depth of relevant solution providers has swiftly accelerated during the past several years to meet advertisers' needs for matching consumers and measuring ad impact across various consumer touch points. The rapid evolution and simultaneous expansion and consolidation of participants in this ecosystem present marketers with a confusing, overlapping collection of companies to navigate. Because of the requirements of navigating capabilities and needing to evaluate competencies, the marketplace confusion is not surprising.

After talking to numerous vendors, users of their services, and several industry experts, CIMM recommends that individuals attempting to navigate the vendor landscape be aware of several broad categories of service suppliers:

1. **Data onboarding:** These companies offer solutions that link traditional customer data (often, but not always, offline information only) with additional data and processing capabilities that assist in creating cross-platform identity matches. In many cases, these solutions either have their own cross-device solutions or work with companies that do. Examples of companies in this space are LiveRamp and Neustar.
2. **Deterministic solutions**: Are companies that provide access to digital audiences at scale via log-in data. Examples of companies in this space are Facebook, Google, AOL Verizon (via its Precision Market Insights solution), and Viant.
3. **Data enrichment providers (DEPs):** Are providers possessing consumer demographic and transactional data sets that can be integrated with other data for a richer picture of consumers. Examples include Experian, Acxiom, Epsilon, Datalogix, and Polk.
4. **Cross-device/cross-platform solution providers:** Companies that triangulate data across the myriad ways platforms, publishers, and ad tech companies try to identify Internet users across smartphones, tablets, and desktop computers as well as offline devices such as digital set-top boxes and smart TVs. Broadly speaking, there are two categories of such vendors:
   a. **End-to-end solutions**: Several companies include cross-device/cross-platform offerings as part of a comprehensive range of relevant solutions including data onboarding and enrichment as well as data matching and validation, both via proprietary assets and working with other third-party vendors. Examples include Oracle and Conversant.
   b. **Probabilistic graph providers:** Solutions used primarily for their ability to correctly match consumers across digital platforms. Examples include TapAd, Drawbridge, Screen6, and Crosswise.
5. **Ad tech companies**: These are primarily digital ad-serving firms, demand supply platforms (DMPs), sales supply platforms (SSPs), and exchanges that are intermediaries for the digital ad-serving process. Examples include Turn, Doubleclick (Google), and Rocketfuel.

It is important to note again how much potential overlap there can be among providers in this space. For example, earlier this year, end-to-end solutions provider Oracle bought cross-device tech company Crosswise for its Data Cloud solution (for a reported $50 million), but it continues to work with probabilistic graph providers Tap Ad (now owned

by Telnor) and Drawbridge to power its Oracle Data Cloud products. Drawbridge, by way of example, offers both a cross-device graph and a cross-device platform that help in targeting advertisements based on their matches.

The Bottom Line—Vendor Selection

To help guide CIMM members in navigating this sea of options, CIMM strongly suggests the following four-step interaction strategy:

a. **Review** internal data processes and assets. Taking inventory both of consumer data on hand and of the companies currently involved in enhancing or managing it is a key first step, as many of these vendors already possess cross-device capabilities. Often, organizations have sets of consumer data that are highly disaggregated and sit across the organization.

b. **Define** them in the ecosystem. Ensure you have a firm understanding of solutions capabilities across the range of key activities required to make cross-device/cross-platform matching successful. Often, a key decision here will be whether to go with an **end-to-end solution** or create a best-of-breed system on your own.

c. **Understand** graphs. No two graphs are the same, and all have individual strengths and weaknesses. This applies to both their accuracy and their relative strengths around particular types of consumers or data sets. For example, Lotame's solution determines relationships that exist between billions of PII-free signals flowing from desktops, smartphones, and tablets, but not digital set-top boxes.

d. **Test** accuracy frequently. Marketers must ask vendors how and what they are matching and never take a number at face value. For example, in calculating the match rate, what is the precise relationship of numerator and denominator to each other? How does a given match rate percentage correspond to the total universe of target consumers?

7 - Vendor Selection
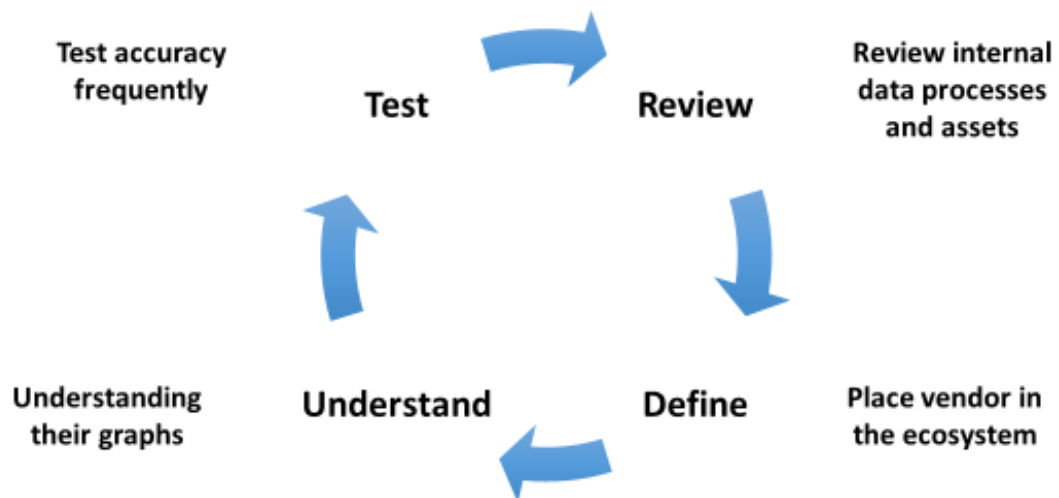
VII. Key Finding: Managing Internal Challenges

Although most of the challenges we discovered in our research were in the realm of dealing with third parties, there is good evidence in many cases that within the enterprise itself, a host of issues are impeding the progress of cross-device matching. "Many people talk the talk of Identity-Based Marketing but do not walk the walk," noted one single-source firm CEO, adding, "too many marketers are taking a ready-fire-aim approach here, running into the market to test the offerings with ill-defined objectives and poor alignment in their own houses to support such projects in a meaningful way." Among the most common issues we heard mentioned were:

a. Restructuring organizations to accommodate the practices associated with Identity-Based Marketing.
b. Setting up the data and technology needed to feed such systems.
c. Dealing with data quality concerns associated with fraud, viewability and even cross-device and online-to-offline data share; and
d. Not asking enough questions to ensure vendors are adhering to privacy-compliant methods of linking identity (e.g., third party blind matching, hashed IDs, etc.).

The Bottom Line—Organizing for Success

There are several practical steps marketers can take to determine how well-suited their own organizations are to engaging in Identity-Based Marketing, starting with an honest assessment, from internal POVs, of the company's position. **"**There are always other things to do," noted James Collins, senior vice president and general manager at Rakuten Attribution (formerly DC Storm). "There's always a logistics and supply chain to improve, there's a product to improve and a website to re-platform. There's a whole bunch of other stuff that will get in your way. Companies will always have opportunities to spend money on something else." Next, marketers must force internal alignment and agreement that the company is pursuing Identity-Based Marketing strategy and come up with a unified view of the customer that can be communicated across the organization, fostering concurrent internal alignment. Brands proceeding forward should start with setting reasonable expectations regarding how much time is needed for both of the above decisions to be reached and implemented; such processes typically take anywhere from 12 to 18 months. During this time, fluid approaches that embrace change management will determine success or failure.

VIII. Key Finding: The Third Rail of Privacy: PII

One issue highlighted by all participants was the ever-shifting landscape of issues around PII and consumer privacy concerns.

Currently, data used for the express purpose of distinguishing individual identity are clearly classified as PII under the definition used by the National Institute of Standards

and Technology, while data points like cookies are considered non-PII because they are traits shared by many people. The gray area here has always been the caveat that non-PII can "potentially" become PII, because they may be combined with other personal information to identify an individual.

"In many ways, the success of probabilistic solutions has drawn a great deal more attention and scrutiny to Non-PII," noted one solution provider we talked to, adding, "by producing highly accurate matches, we have, in effect, proven that non-PII can be combined successfully to identify an individual, and that genie cannot be put back in the bottle."

While speaking recently to Network Advertising Initiative members, Jessica Rich, the FTC's Consumer Protection Bureau Chief, focused attention on these practices while describing the FTC's position on persistent identifiers and privacy. In a follow-up blog post, Rich noted, "We regard data as 'personally identifiable,' and thus warranting privacy protections, when it can be reasonably linked to a particular person, computer, or device. In many cases, persistent identifiers such as device identifiers, MAC addresses, static IP addresses, or cookies meet this test."

"This notion that, in effect, any data is personal data, that an IP address is personal data, et cetera, is not new, and has been the general thinking if not literal law in the EU for the past four years," notes well-known digital privacy expert Alan Chappell of Chappell & Associates, adding, "It increasingly looks like we could be moving more towards a fairly broad definition of PII in the United States as well. It's becoming harder and harder for anybody to say, 'Oh, it's just the Europeans,' or, 'It's just the FTC.' It just seems like the trend is to have an extremely broad definition of personal data."

Deploying a broader definition of PII goes hand in hand with the issues around the current mechanisms for consumer consent, specifically methods of opting out on data collection and resultant targeting. Justin Bookman, policy director of the FTC's Office of Technology Research, speaking at the FTC's Cross-Device Tracking Workshop, noted that "It's fair to say this area is evolving rapidly and may be… challenging traditional consumer expectations about their privacy…. It's really hard to determine objectively, from the end user point of view, when cross-device tracking is going on… [And] that raises the question: How much transparency should there be? What do consumers expect? Do they want to be overloaded with information? If cross-device tracking is going on, what should consumers be told and how?"

8 - Five Principle Takeaways from the FTC's Cross-Device Tracking Workshop

1. Cross device tracking can provide benefits, including reducing the number of times a consumer sees an ad (preventing ad fatigue).
2. Greater transparency, choice and consumer education are needed.
3. Additional attention should be paid to the consumer experience as technology continues to evolve.
4. Room remains for industry innovation, including developing controls to increase consumer awareness of tracking.
5. Companies must be mindful of the representations they make. In particular, they should be careful in accurately describing what it means to opt out. The failure to do so can be construed as deceptive and could lead to FTC action.

See https://goo.gl/9cNMXb for additional details

9

"You're going to see a couple of companies get dinged on this, unfortunately, in the next six months," said Chappell "Perceived risks to privacy and security, even if they never happen, can undercut consumer confidence necessary for the technologies to meet their full potential and could well hinder adoption."

The Bottom Line—Privacy

The intersection of increased government scrutiny and heightened consumer concerns around privacy require marketers approaching identity-based marketing to have a deep understanding of the data sources they are working with and how they do or do not protect consumer privacy. Key factors to review here include:

1. **Ingredients:** Is PII involved in any part of your matching process? If so, how is PII protected? How is it sourced? Did you take steps to ensure that the vendor's collection complied with privacy laws and their privacy practices?
2. **Policies:** Review in depth any potential partner's privacy policies and ensure that if they do use PII you have a full understanding of how their policies are implemented and how they will honor your privacy commitments to consumers
3. **Execution**: Most vendors that collect PII typically have very detailed systems to manage this data. As a best practice these should include:
   a. Ability for consumers to opt out of the data set in question.
   b. A Chief Privacy Officer (certainly at bigger companies) or at minimum, a person designated with responsibility for privacy.
   c. Accurate explanations of data usage and policies in any relevant terms of service (TOS) and privacy policy presented to consumers.
   d. Numerous third-party matching safeguards such as hexed files.
4. **Self-Regulation:** Ensure your vendor participates in the fullest range of industry self-regulation efforts, such as those conducted by Network Advertising Initiative (NAI), the Interactive Advertising Bureau (IAB), AdChoices and AppChoices,

created by the Digital Advertising Alliance (DAA), TRUSTe and the Direct Marketing Association (DMA) XDID initiative, which is developing an RFI for buyers and sellers to help clarify the purchasing process for identity-matching services. Additionally, compliance with guidance from regulators like the FTC (whether published or inferred from public statements) is key here as well.

9 - Factors to Review: Privacy

FINAL RECOMMENDATIONS

In the process of researching this report, CIMM identified several key areas that require additional focus by and across industry bodies. These include:

- Third-party validation of key measures such as match rates and accuracy.
- Industry-level education in terms of best practices for identity matching and assessment of data quality.
- Stronger clarity and standards about protecting consumer privacy.
- Industry standardization definition of key measures and agreement on how they should be calculated.

ACKNOWLEDGEMENTS

APPENDIX A: A GLOSSARY OF KEY TERMS

Accuracy: A measure of correctly identified **matches** and **non-matches** out of the total predicted.

Cross-Device Attribution: Traditional single-device attribution wrongly assumes all message exposure and activity occur on the same device. Cross-device attribution, however, properly assigns value to each device on the path to purchase, enabling marketers to better understand their consumers' behavior and more intelligently invest their budgets.

Cross-Device Marketing: Cross-device marketing allows marketers to deliver brand messages to consumers at the right time in the right format on the right device. The underlying foundation of cross-device marketing is a technology that links together various identifiers associated with a digital consumer. This linking forms a holistic profile of an individual, which enables marketers to address and understand actual people instead of devices.

Device Graphs: A device graph describes how different devices relate to each other by mapping all the devices, IDs, and associated data back to one unique user or household.

Deterministic Data: Cross-device marketing using deterministic data creates links between devices with a 100% level of confidence using log-in or subscriber data. For example, a user will log in to a streaming music service with the same log-in on their desktop and mobile app. This will confirm that those devices belong to the same individual. Although highly accurate, this type of data is limited in scale.

Household Data: Household data refers to data collected from non-personal devices that are shared within households, such as a television set or a desktop computer for a family. Household data has broad scale but less precision at an individual consumer level, given the shared nature of the targeted devices in question. Traditionally, this is a common way television advertising has been bought—with an entire household in mind.

Identity-Based Marketing: Often referred to as **People-Based Marketing**, identity-based marketing is focused on the opportunity to reach consumers in a highly holistic, targeted, interactive mode across the widest range of activities and channels.

Match Rate: The percentage of unique records in the client's data set that can be matched to an identifier in the solution provider's database.

Precision: In a probabilistic model, precision estimates how well the model correctly identifies **matches,** taking into account that some of those matches will be in error.

Probabilistic Data: Probabilistic data uses algorithms and patterns to connect devices with high levels of confidence. For example, if two devices always move together to the same locations, at the same time every day, they likely belong to the same

individual. High-quality providers will use a truth set to validate the accuracy of their models. This type of data is generally reliable and scalable.

Reach (or Scale): Reach is an estimate of the breadth of matches correctly identified; that is, how many cross-device identities or devices the solution provider can correctly match in the context of a given use-case and desired level of precision. Some cross-device solution providers use the term "recall" to represent reach; these solution providers will publish their recall rates relative to a level of precision. Recall, reach, and scale are synonymous.

Truth Set: Refers to a data set that has been verified as 100% accurate. Often used as part of the process by which data matches are validated.